

Combining Field and Imaging Spectroscopy to Map Soil Organic Carbon in a Semiarid Environment

Anita D. Bayer, Martin Bachmann, Derek Rogge, Andreas Müller, and Hermann Kaufmann

Abstract—Semiarid regions are especially vulnerable to climate change and human-induced land-use changes and are of major importance in the context of necessary carbon sequestration and ongoing land degradation. Topsoil properties, such as soil carbon content, provide valuable indicators to these processes, and can be mapped using imaging spectroscopy (IS). In semiarid regions, this poses difficulties because models are needed that can cope with varying land surface and soil conditions, consider a partial vegetation coverage, and deal with usually low soil organic carbon (SOC) contents. We present an approach that aims at addressing these difficulties by using a combination of field and IS to map SOC in an extensively used semiarid ecosystem. In hyperspectral imagery of the HyMap sensor, the influence of nonsoil materials, i.e., vegetation, on the spectral signature of soil dominated image pixels was reduced and a residual soil signature was calculated. The proposed approach allowed this procedure up to a vegetation coverage of 40% clearly extending the mapping capability. SOC quantities are predicted by applying a spectral feature-based SOC prediction model to image data of residual soil spectra. With this approach, we could significantly increase the spatial extent for which SOC could be predicted with a minimal influence of a vegetation signal compared to previous approaches where the considered area was limited to a maximum of, e.g., 10% vegetation coverage. As a regional example, the approach was applied to a 320 km² area in the Albany Thicket Biome, South Africa, where land cover and land-use changes have occurred due to decades of unsustainable land management. In the generated maps, spatial SOC patterns were interpreted and linked to geomorphic features and land surface processes, i.e., areas of soil erosion. It was found that the chosen approach supported the extraction of soil-related spectral image information in the semiarid region with highly varying land cover. However, the quantitative prediction of SOC contents revealed a lack in absolute accuracy.

Index Terms—Imaging spectroscopy (IS), land degradation, linear spectral unmixing, multiple linear regression analysis, soil organic carbon.

Manuscript received March 15, 2016; revised June 22, 2016; accepted June 24, 2016. Date of publication August 3, 2016; date of current version September 30, 2016. This work was supported by the Helmholtz EOS Network, a collaboration of German Helmholtz research centers. The PRESENCE Network (a Living Lands initiative), Mike Powell (Rhode University, South Africa), and the University of Stellenbosch, South Africa, supported the intensive field sampling. (Corresponding author: Anita D. Bayer.)

A. D. Bayer was with German Aerospace Center (DLR), German Remote Sensing Data Center, Department of Land Surface, Wessling 82234, Germany. She is now with the Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Atmospheric Environmental Research, Garmisch-Partenkirchen 82467, Germany (e-mail: anita.bayer@kit.edu).

M. Bachmann, D. Rogge, and A. Müller are with the Department of Land Surface, German Aerospace Center (DLR), German Remote Sensing Data Center, Wessling 82234, Germany (e-mail: martin.bachmann@dlr.de; derek.rogge@dlr.de; andreas.mueller@dlr.de).

H. Kaufmann was with the Department of Geodesy and Remote Sensing, Remote Sensing Section, German Research Centre for Geosciences (GFZ), Potsdam 14473, Germany.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2016.2585674

I. INTRODUCTION

SEMIARID regions are especially vulnerable to climate change and human-induced land-use changes as these processes often lead to a significant decrease in available water resources that can destabilize these regions. This is accompanied by a decrease in the ecosystem carbon reservoirs in vegetation and soil (e.g., [1], [2]). Optical remote sensing and in particular imaging spectroscopy (IS) has shown the potential to assess several soil properties, among others soil organic carbon (SOC), over large areas, taking advantage of diagnostic absorption characteristics inherent in the reflectance spectrum of a soil (e.g., [3]–[5]). In this regard, the Intergovernmental Panel on Climate Change (IPCC) [6] identified “improved remote sensing technologies for analysis of soil carbon sequestration potential and land-use change” as one of the key methods underpinning sectoral mitigation technologies for the coming decades. However, large-area analyses of soil constituents and carbon stocks in extensively used semiarid areas using IS poses difficulties because models are needed that can cope with varying land surface and soil conditions, that consider partial vegetation coverage, and that are able to predict SOC concentrations that may be significantly lower than in agricultural regions. Consequently, soil constituents and soil carbon stocks are assessed at small plots in cost- and time-consuming field campaigns.

Nevertheless, reflectance spectroscopy was shown to be a promising approach. This technique is increasingly used in laboratory and field environments to predict various soil constituents based on their diagnostic spectral features and statistical regression approaches (e.g., [3], [7]–[9], and references in [10]). A special interest was given to SOC due to its importance for soil stability and fertility, together with its contribution to the global carbon cycle (e.g., [11], [12]). The most frequently used technique to establish soil prediction models is partial least squares regression analysis (PLS, e.g., [4], [5], [13]–[15]), which proved particularly useful when the number of independent variables exceeds the experimental observations. However, the technique faces drawbacks in terms of significance and transferability resulting from the high degree of adaptation to the applied data making it sensitive to dataset-inherent properties such as noise. Attempts to predict samples from another study site or dataset with a PLS model have shown relatively low validation results (e.g., [16]–[18]), which limits the application of PLS models derived from local observations to a regional scale with variant conditions, as they are present in semiarid regions.

Another possibility to establish soil prediction models is the direct application of known spectral absorption features providing an immediate physical basis or empirical approaches such as

band indices that are based on such features. This is often seen as more robust, hence, allowing for a better transferability to regions of similar environmental conditions [19]. The direct application and parameterization of spectral features in general is limited due to the influence of the overlapping of many spectral features, and additional effects of nonsignificant factors (e.g., surface condition, sensor noise). Nonetheless, there remain diagnostic spectral features which can be used for a quantification of soil constituents, amongst others SOC (see, e.g., [11], [20]). Regression models built on one or a few spectral features are often not sufficiently robust for a practical application to a wide variety of soils [5].

With increasing frequency in recent years, airborne imaging spectrometer data have been used for predictions of soil constituents (e.g., [3], [4], [13], [21]–[23]). Besides the IS technique necessitating atmospheric and geometric correction of the data, the raster-like recording of spectral information introduced the mixing of spectral signatures of the surface materials present in a pixel's area which disturbed the relationship between soil constituents and reflectance by the contribution of other materials, mostly vegetation, to image pixel spectra [19]. A traditional approach to deal with this influence was to limit the image information before the analysis to soil-dominated pixels that comprise only a small amount of vegetation, often determined using vegetation indices or spectral mixture analysis (e.g. [3], [24]). An influence of the remaining vegetation on soil analysis was accepted. This is especially significant for SOC predictions because a number of spectral features of soil carbon occur in similar spectral ranges as those of vegetation components, especially found in dry vegetation [25]. Soil constituents for the excluded areas were eventually interpolated afterward from the nonvegetated areas.

As an alternative, methods that decompose spectral mixtures within image pixels are used to extract the spectral signal of the soil before an analysis of soil constituents. For instance, [26] used unmixing ground cover fractions to remove the influence of maize from the spectra of partly vegetated agricultural fields in Belgium before a prediction of SOC from AHS-160 data using PLS regression techniques. With this technique, they obtained a validated ratio of performance to deviation (RPD), RPD_{Val} , of 1.50 in SOC predictions and showed that the error in prediction rose with vegetation coverage and that this could partly be overcome by the application of what they called residual spectral unmixing. With the objective of clay content estimations, [27] applied blind source separation to resolve spectral mixtures in southern France. The approach has the advantage that spectral signatures can be resolved without any prior knowledge. They found, same as [26], that using PLS regression analysis no accurate predictions from mixed HyMap pixels could be achieved beyond a vegetation cover of 10%. Using blind source separation, this could be improved to a vegetation proportion of 30%. More recently, Franceschini *et al.* [28] explored the selection of optimal spectral datasets for the prediction of several soil properties and investigated the effect of decreasing bare soil fraction. All three applications were focused on agricultural areas where spectral

mixtures comprised of mostly two endmembers and could be well described and modeled. So far, this has not been tested in natural areas, where spectral mixtures are more complex due to many spatially variant surface materials.

In the given context, this research study aims at quantifying SOC in a semiarid area using field and IS. Difficulties arising with SOC predictions in extensively used semiarid areas are addressed in a multistep methodology:

- 1) In the hyperspectral imagery, pixel spectral signatures that contain a mixture of soil and vegetation are deconvolved into their constituent components using spectral mixture analysis.
- 2) For soil dominated pixels, the residual soil signature is calculated.
- 3) SOC quantities are predicted from the processed hyperspectral imagery using a spectral feature-based SOC-prediction model (the development of the SOC-prediction model was published in [29]).

Thematic maps of SOC distribution are evaluated for their potential to detect erosion affected areas as effects of land-use changes. With these aspects, the main focus of this study lies in the extension of existing approaches leading toward the mapping of SOC contents to facilitate their application in semiarid environments. The methodology is tested in an area of 320 km² within the Albany Thicket Biome in the Eastern Cape Province of South Africa. It provides a regional example for an extensively used semiarid ecosystem that currently undergoes land-use changes due to unadapted management practices, and that is likely to face climate change induced land-use changes in the future [30]. The accuracy of achieved SOC predictions is assessed and discussed in the given context along with the achievements of all processing steps.

II. STUDY AREA AND DATA COLLECTION

A. Land-Use Changes in the Albany Thicket Biome

The study area is located in the Albany Thicket vegetation type covering the semiarid valleys of the Eastern Cape Province of South Africa [31]. In more than 70% of the area, the originally dense vegetation is “moderately” to “severely” degraded [32] as a result of intensive goat farming since the early 1900s. The ecosystem transformed to an open savannah-like system under the loss of biodiversity and ecosystem carbon stocks [31]. Current land cover ranges from bare surfaces to dense shrub vegetation (average height 2–3 m). The region has been identified as highly important for carbon sequestration projects due to the unique characteristics of the local vegetation type in its original state to sequester comparably high amounts of carbon in biomass and soils [33]. Within the Albany Thicket region, an area of approximately 320 km² consisting of two parts is selected as study area. The area is representative on regional scale and serves as an example for larger scales (e.g., biome-wide assessments). A detailed description of the study area, its climate, terrain elevation, geology, soils, and a map can be found in [29].

B. Field Sampling and Chemical Analysis

We use the data from two field surveys conducted in the study area in June and September/October 2009 where in total 163 natural, nonagricultural sites on sandy and loamy soils with generally low organic carbon contents ([29, Fig. 3(a)]) were sampled for ground truth. We use the SOC contents determined for 163 sampled sites and field spectra of bare soils (125 sites) resampled to HyMap's spectral resolution as they are soils were described in [29]. Details on the collection of field data, including soil field spectra and soil samples can be accessed there.

C. Airborne Hyperspectral Imagery

Airborne HyMap imaging spectrometer data were acquired for the 320 km² study area on 14th October 2009. HyMap scans the 440–2500 nm spectral region with 126 bands, resulting in average spectral resolutions between 13 and 17 nm depending on the wavelength range (see [34]). Northern and southern sections were recorded separately in three flight lines each with a length of 57.2 km for the northern and 22.4 km for the southern section. Ground resolution of the imagery is 3.3 m. Orthorectification is performed using the PARGE procedure for geocorrection of airborne data [35], with the additional usage of ground control points selected based on Google Earth and a digital elevation model available from the SRTM mission. Typical accuracy of this procedure is better than ± 3 pixels (approximately 10 m). At-sensor radiance data are corrected for atmospheric and terrain illumination effects using the ATCOR-4 procedure [36]. This includes inflight calibration, where, due to the lack of large invariant artificial targets in both parts of the study area, a large homogenous bare soil area is used for the correlation of ground and airborne measurements. Eleven HyMap bands are classified as reduced quality and removed (427–442 nm, 889 nm 1956–2050 nm, 2471–2486 nm), resulting in 115 bands. Small remaining spectral artifacts are locally smoothed using a low pass filter with a filter width of three bands. The three individual flight lines of each section are leveled to each other correcting for brightness and radiometric differences across the flight lines following an approach described by [37]. Mosaicked hyperspectral imagery is spatially filtered with an iterative adaptive spatial filtering approach [38], which minimizes random noise and local intraclass spectral variability on a pixel-to-pixel basis. Spectral contrasts between different materials and, therefore, regional differences are retained. This supports the identification of endmembers and the unmixing process. Artificial objects and rural settlements are masked to limit subsequent analysis to natural surface materials.

III. METHODS

A. Extraction of Soil Signatures From Mixed Pixels

The hyperspectral imagery of the HyMap sensor reflects a large spatial heterogeneity of the study area that is caused by the patchiness of natural vegetation inherent to most semiarid landscapes. Pixel spectral mixtures are resolved and per-pixel soil signatures are approximated in three steps: 1) derivation of spectra of pure surface materials (endmembers, EM); 2) linear

spectral mixture analysis; and, 3) calculation of residual soil spectra (see Fig. 1).

Potential endmember spectra are derived directly from the image data using the Spatial–Spectral Endmember Extraction tool (SSEE) described by [43]. In addition to detecting endmembers based on spectral characteristics alone, the algorithm includes spatial information by analyzing a scene in parts, which allows for an identification of endmembers that have a low spectral contrast with each other, but occur spatially independent. From the pool of SSEE endmembers, a representative number of endmembers is manually selected depending on the present variance of the endmembers, and they are assigned to the land cover classes photosynthetic active (PV) and nonphotosynthetic active vegetation (NPV) and bare soil (BS). One endmember set is prepared for the northern (235 km²) and one for the southern section of the study area (92 km²) since the two areas differ in their characteristics.

Spectral mixture analysis is used to deconvolve image pixel spectra into fractional abundances of the endmembers [44], [45]. Recent algorithms apply a linear approach, in which the reflectance of an image pixel (ρ_S) is assumed to be a linear combination of each endmember's reflectance spectrum (ρ_{EM}) and the fraction of it within the mixture (f_{EM}). The following equation is solved by a least squares approach, minimizing the error term (E):

$$\rho_S = \sum_{i=1}^n \rho_{EMi} \cdot f_{EMi} + E. \quad (1)$$

The unmixing root mean square error (RMSE) is a measure of the goodness of fit of pixel spectrum and modeled spectrum. Unmixing is performed using the Iterative Spectral Mixture Approach (ISMA, see [46]). ISMA applies an iterative procedure to determine which endmembers are best suited to resolve the mixture of a given pixel and to solve the linear mixture problem in unmixing. A uniform reflectance shade endmember is also included. The maximum number of endmembers to be used for each pixel is set to 5.

Spectral unmixing results are assessed using ground cover fractions assessed in the field. This comparison is flawed by a number of factors, e.g., the subjective nature of personal assessments, the time difference between field assessments and image data acquisition, inaccuracies in sample plot locations, and the dependency of unscaled cover fractions on endmember brightness. As the accuracy in georeferencing is about 3 pixel RMSE, the validation based on matching single pixels with ground references will likely be unsuccessful. To reduce the influence of the geolocation error on the ground cover validation, for every ground cover measurement taken in the field, a neighborhood of 3×3 pixels around this coordinate is selected in the image and the closest match between the image-derived ground cover fractions within this window and the ground reference is chosen for validation. This is done because for areas with highly heterogeneous vegetation cover, already small discrepancies in geolocation might cause a big difference in cover which could hinder a reasonable comparison.

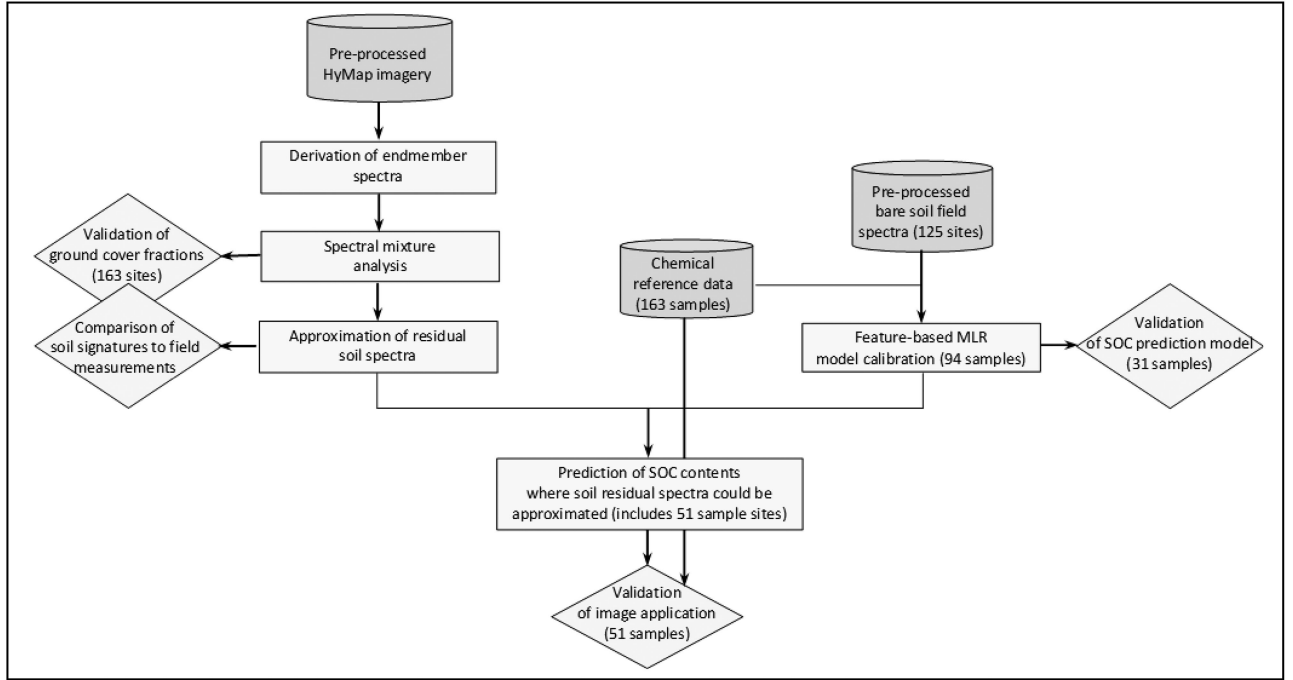


Fig. 1. Processing workflow for the prediction of SOC from image data.

In a last step, the estimates of ground cover fractions retained by spectral unmixing are used to remove the spectral influence of vegetation on a nonpure soil spectrum. Thereby, the soil signature contained in a mixed pixel is approximated. In the following, this is referred to as residual soil spectrum (RSS), a term that was first used by [26]. RSS is described by the inversion of the linear mixture model by applying the endmember spectra of the classes PV and NPV contributing to the spectral mixture of the pixel and their cover fractions

$$\rho_{RSS} = \rho_S - \rho_{EM1} \cdot f_{EM1} - \dots - \rho_{EMn} \cdot f_{EMn} \quad (2)$$

with the reflectance ρ_{RSS} of the RSS, the mixed pixel spectrum ρ_S recorded by the HyMap sensor, the spectrum ρ_{EM} of the n individual endmembers of PV and NPV used to unmix the specific pixel and their unscaled cover fractions f_{EM} . RSS is supposed to include the spectral signal originating from the soil fraction and any signals which are not explained within spectral mixture analysis, e.g., noise [as part of E in (1)]. The extracted RSS is scaled by the summed fractions of the nonsoil endmembers (n) identified in the pixel to approximate a full “endmember” spectrum

$$\rho_{RSS_{sc1}} = \rho_{RSS} \cdot \left(1 - \sum_{i=1}^n f_{EMi}\right). \quad (3)$$

The spectral mixture characteristics in natural environments hundreds of square kilometers in size can be locally very complex due to land cover changes coupled with topography, land-use, etc. This results in a global endmember set which is not necessarily suitable for the unmixing of every pixel. Thus, additional criteria are needed to filter out pixels where RSS might be prone to errors owing to the predominance of a land cover class

that is not BS or because of insufficient unmixing. Those pixels are masked for further analysis. The following quality requirements are set up: 1) minimal soil fractional cover, 2) maximal fractional cover of PV and NPV, 3) maximum unmixing RMSE, 4) range for the sum of the cover fractions around 1, and 5) minimum mean reflectance of the RSS to filter out dark pixels where unexplained signal such as noise highly influences RSS. It is assumed that the soil signature included in the spectral mixture of a pixel can be approximated with a high accuracy if these requirements are fulfilled. This procedure results in a reflectance image of the BS surface. To evaluate the accuracy of residual soil spectra, they are compared to corresponding field reference spectra. In order to highlight the differences in spectral shape, spectra are normalized to mean reflectance.

B. Feature-Based SOC Prediction Model

For the prediction of SOC contents from spectra, we use a feature-based MLR model that applies the parameters (e.g., absorption band depth, width, slope in a given wavelength range) of diagnostic spectral absorption features and characteristic changes in reflectance in the VIS and the NIR to SWIR range. The set-up of the SOC calibration model and its test set validation is described in detail as “Approach A: multiple linear regression of spectral feature parameters” in [29]. The approach benefits from the collinearity of several diagnostic spectral features and their unique parameters, which in practice is weakened by subtle variations in the makeup of the material (e.g., influence of other substances or compositional variations) and potentially variability related to image acquisition (e.g., sensor-dependent noise, illumination conditions) that impact these variables individually and in a nonlinear fashion. Therefore, spectral variables

TABLE I
CORRELATION MATRIX (R^2) OF SPECTRAL VARIABLES OF 125 SPECTRA

	r_{SWIR1}	s_{VIS}	r_{VIS}	$d_{\max 1730}$	AS_{2330}	s_{SWIR1}	A_{1730}	$d_{\max 2330}$	$\lambda_{d\max 2330}$	$\lambda_{d\max 1730}$	w_{1730}	w_{2330}
r_{SWIR1}	1											
s_{VIS}	0.70	1										
r_{VIS}	0.60	0.52	1									
$d_{\max 1730}$	0.00	-0.03	-0.06	1								
AS_{2330}	0.00	0.01	-0.02	0.14	1							
s_{SWIR1}	0.28	0.06	0.04	0.00	-0.03	1						
A_{1730}	0.00	-0.02	-0.04	0.95	0.13	0.00	1					
$d_{\max 2330}$	-0.04	-0.09	-0.01	-0.00	-0.05	0.01	-0.00	1				
$\lambda_{d\max 2330}$	0.01	0.01	-0.00	0.08	0.53	-0.00	0.11	-0.03	1			
$\lambda_{d\max 1730}$	0.01	0.00	0.01	0.06	-0.00	0.01	0.03	0.01	0.02	1		
w_{1730}	0.03	-0.00	-0.00	0.61	0.04	0.00	0.59	0.00	0.01	0.29	1	
w_{2330}	-0.00	-0.04	-0.01	0.01	-0.01	0.02	0.00	0.18	-0.05	0.00	0.00	1

show a loosened connection between each other (see Table I) and can be used for regression analysis. The model developed for field spectra of BSs in HyMap's spectral resolution is used. Note that only 115 HyMap bands are used in this study compared to 116 in [29]. This did not affect model calibration or validation accuracies.

BS field spectra are used so as to exclude the spectral variability introduced by small stones (see [39]). HyMap spectra recorded at the sample locations are not used, because due to the small scale heterogeneity of the South African study area, nearly every HyMap pixel is comprised of a mixture of several surface materials. Even though a method is proposed in this study to derive the soil signature of image pixels by the use of spectral unmixing, it is decided not to include the large influence of spectral unmixing in the calibration procedure. Several processing steps are taken to minimize the spectral differences between field and HyMap acquisitions to allow the transfer of the calibration model built based on extensive field information to HyMap scanner data (see [40]). This includes radiometric vicarious calibration of image data using field reference spectra to minimize the difference in ground reflectance after atmospheric correction, spectral resampling of field spectra, and extraction of soil signatures from HyMap spectra before SOC prediction taking into account most effects of increased spatial averaging. The transfer of soil prediction models calibrated based on field spectra to airborne scanner data has been demonstrated before, considering that appropriate corrections applied that support this transfer (e.g., [41]). For the HyMap sensor, field and laboratory spectra calibrated PLS models were applied with satisfactory accuracy by, e.g., [24] and [42].

C. Prediction and Validation of Soil Constituents From Hyperspectral Imagery

SOC contents are predicted by applying the SOC prediction relationship to the hyperspectral image data of residual soil signatures. Image spectra are parameterized and spectral variables are standardized in the same way as for the field

spectra during calibration (see [29]). SOC contents are calculated from the spectral variables according to the regression relationship, which is done for each pixel, a RSS is calculated for. Pixel values exceeding the calibration range are detected and masked.

The performance of the prediction models is evaluated based on coefficient of determination (R^2) for predicted versus measured compositions, root mean squared error (RMSE), and RPD grouped to three classes according to [16]. To validate model derived SOC with *in situ* field measurements, all sample points are used, for which an SOC content is predicted for the specific pixel in the imagery comprising the GPS sample location. This results in 51 samples. Although this is a low number compared to the 163 originally sampled sites (see Fig. 1), these samples still provide a suitable basis for validation of the multistage methodology. For the remaining 112 sites, a sample was collected in the field, however, an RSS could not be adequately derived from hyperspectral imagery, e.g., because of a too high vegetation coverage within the pixel (see Section 3.2). Measured reference SOC contents are compared to predicted values. For this, in a 3×3 neighborhood around the specific pixel comprising the Universal Transverse Mercator (UTM) coordinates of the sampled site (center pixel) all available predictions were averaged to account for geometric location errors and allow for a SOC content that is representative for the area.

IV. RESULTS

A. Extraction of Soil Signatures From Mixed Pixels

The endmember sets prepared for the northern and southern section of the study area comprise 31 and 21 endmembers, respectively. According to the spectral variability of each class, each set includes more endmembers for green than dry vegetation. With the determination and application of the optimal set of endmembers to be used for the unmixing of each pixel, ISMA ensures a realistic unmixing resulting in low RMSE (average of 2.2%), non-negative abundances, and the sum of the abundances close to one (average sum of 0.69

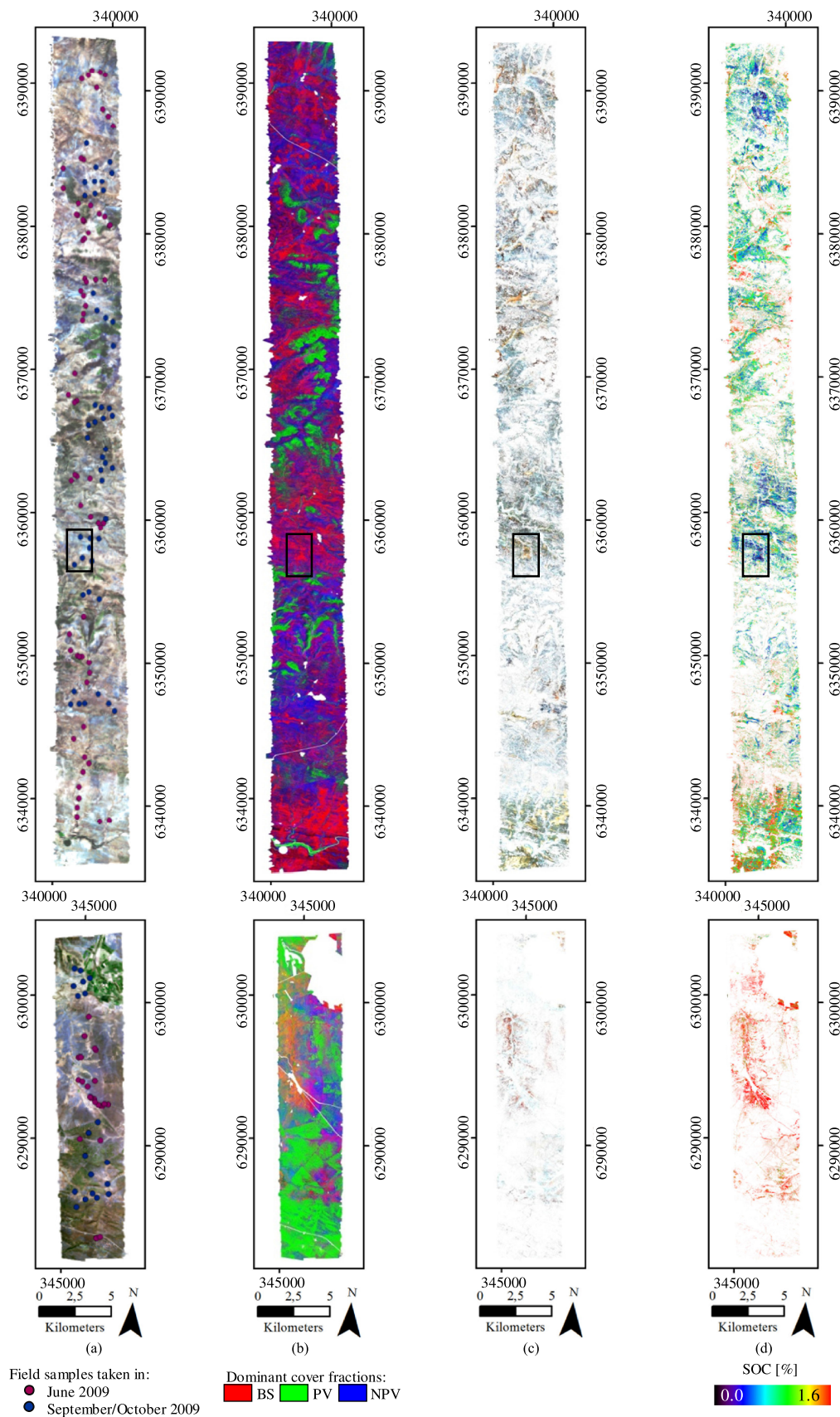


Fig. 2. Application of the multistage methodology in the 320 km² South African study area (see [29] for location). True color of HyMap image data (a), results of spectral unmixing as RGB composite (b), true color image of residual soil spectra (c), and predicted SOC contents (d). White areas are pixels where artificial objects are excluded within preprocessing and where no residual soil spectrum is calculated. Coordinates are given in UTM system (zone 35 South, datum WGS-84).

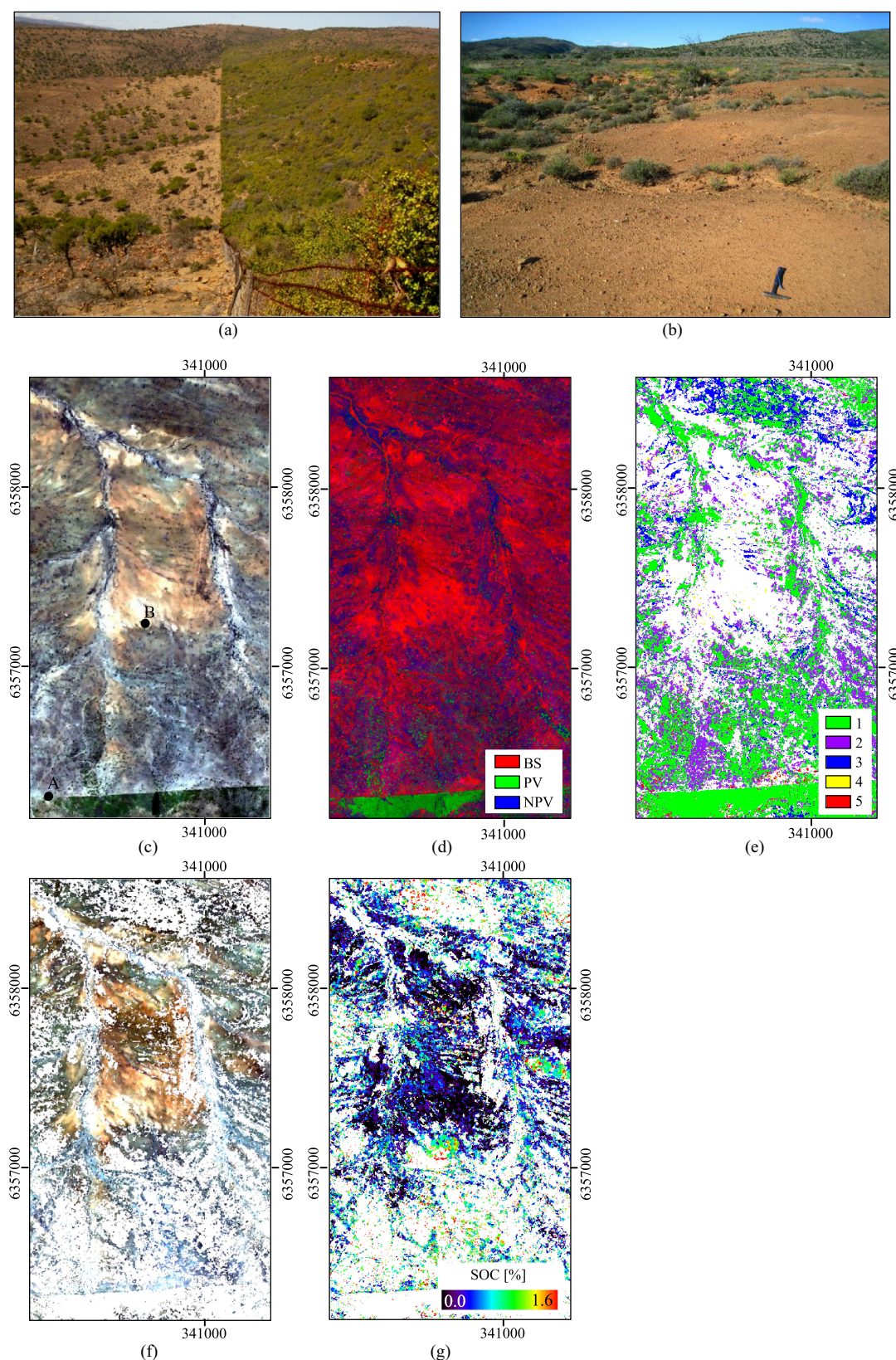


Fig. 3. Prediction of SOC for a selected subset of about 4 km² in the Albany Thicket Biome, South Africa (see Fig. 2 for location). The area shows typical vegetation patterns, with drainage systems dividing it in a bare plain in the center (b, location B in c) and surrounding lowlands. The terrain rises to the south up to a fence line showing the contrast between pristine and degraded vegetation (a, location A in c). True color of HyMap image data (c), results of spectral unmixing as RGB composite (d), pixel exclusion code (e) in the calculation of residual soil spectra (see Table II), true color image of residual soil spectra (f), and predicted SOC contents (g). White areas are pixels where artificial objects are excluded within preprocessing and where no residual soil spectrum is calculated. Coordinates are given in UTM system (zone 35 South, datum WGS-84).

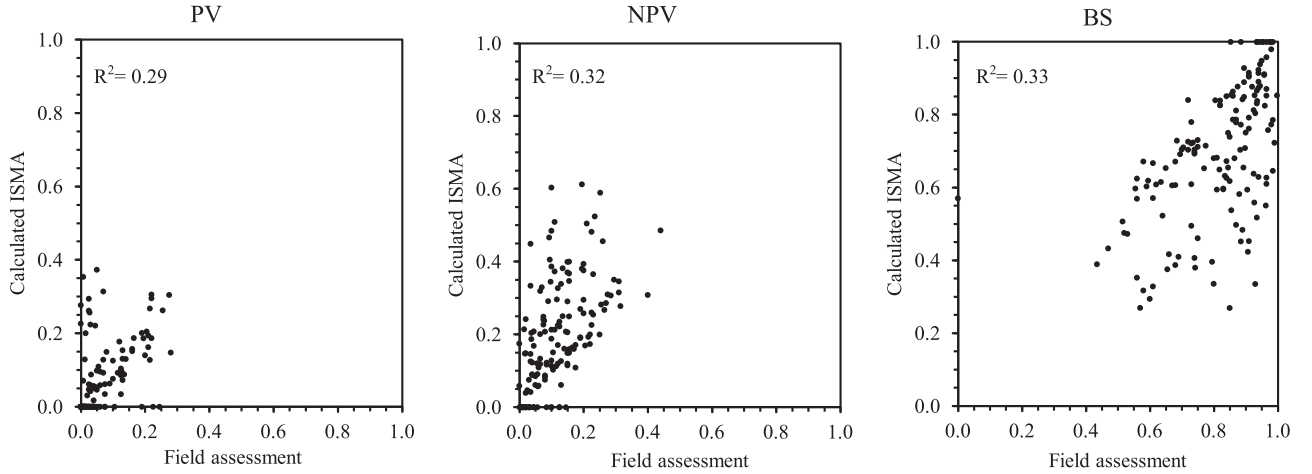


Fig. 4. Comparison of cover fractions of green (PV) and dry vegetation (NPV) and bare soil (BS) from field assessments and calculated with ISMA unmixing (scaled sum-to-one).

TABLE II
CALIBRATION AND TEST SET VALIDATION ACCURACIES OF THE FEATURE-BASED SOC PREDICTION MODEL DEVELOPED BASED ON BARE SOIL FIELD SPECTRA IN HYMAP'S SPECTRAL RESOLUTION

Calibration (94 samples)	R^2_{Cal}	0.81	RMSE _{Cal}	0.43	RPD _{Cal}	2.28	<i>p</i> -value	2.2×10^{-16}
Validation (31 samples)	R^2_{Val}	0.62	RMSE _{Val}	0.43	RPD _{Val}	1.57		-
	Regression coefficient	Influence on regression [%]	Partial correlation coefficients (R^2)	Significance (<i>p</i> -value)				
Constant factor	1.17							
r_{SWIR1}	1.51	30.42	0.10	4.18×10^{-12}				
s_{VIS}	-0.92	-18.57	-0.15	4.04×10^{-12}				
r_{VIS}	-0.59	-12.01	-0.28	4.63×10^{-6}				
$d_{max1730}$	0.46	9.25	0.62	0.12				
AS_{2330}	0.37	7.42	0.49	3.77×10^{-5}				
s_{SWIR1}	-0.34	-6.82	0.002	1.27×10^{-4}				
A_{1730}	-0.21	-4.29	0.55	0.43				
$d_{max2330}$	-0.21	-4.15	-0.30	6.25×10^{-4}				
$\lambda_{dmax2330}$	-0.20	-4.01	0.24	1.57×10^{-2}				
$\lambda_{dmax1730}$	-0.06	-1.21	-0.06	0.42				
w_{1730}	-0.06	-1.20	0.40	0.61				
w_{2330}	-0.03	-0.66	0.12	0.58				
Sum of absolute influences		100.00						

Abbreviations of spectral feature properties: A: area ($A_{left} + A_{right}$), d_{max} : maximum depth, λ_{dmax} : wavelength of d_{max} , w : width; AS: asymmetry factor (A_{left}/A_{right}), r : mean reflectance in interval, s : slope in interval (see, [29], for details). The influence (i) of each spectral variable in the regression relation is given in % of the summed absolute values of all regression coefficients. Negative signed influences indicate negative regression coefficients.

± 0.17). Fig. 2 gives an overview over the entire study area while Fig. 3 shows a subset of the typical landscape in larger detail.

A good correlation between cover fractions calculated by ISMA and field assessments is supported by a number of points (see Fig. 4). However, a substantial amount of points lie apart from the 1:1-line and, therefore, the correlations of the ground cover fractions achieved here do not reach the quality of previous studies where the absolute accuracy of unmixing was determined (e.g., R^2 of 0.74 by [48]). The most likely reason for the deviations is the inaccuracy in geo-correction that, in this heterogeneous area with a high variability on a pixel-to-pixel basis, can lead to huge differences in cover fractions and spectral mixtures as recorded in the image pixels.

When using suitable pixels only, an RSS can be calculated for 26.8% of the image pixels, where the spatial distribution is limited owing to the occurrence of vegetation, shadow, etc. [see Fig. 2(b) and (c)]. The thresholds applied to provide a separation of pixels which are suitable for this procedure based on their spectral unmixing parameters are scene dependent and vary for the northern and southern section of the study area (see Table III, [39]). Residual soil spectra are calculated nearly spatially continuously for soil-dominated areas, whereas they only cover local patches, where vegetation occurs in different intensity such as the southern slopes. Residual soil spectra are calculated up to vegetation coverage of about 40% (combined PV and NPV).

Residual soil spectra are equal to the original pixel spectra if no influence of vegetation is determined in spectral unmixing.

TABLE III
THRESHOLDS AND STATISTICS OF THE CALCULATION OF RESIDUAL SOIL SPECTRA

Subsets	Thresholds							% of pixels ^c					
	Minimal fraction BS	Maximal fraction PV	Maximal fraction NPV	Maximal fraction PV+NPV	Minimal mean reflectance [%]	Maximal RMSE	Sum of cover fractions	Valid Pixels	Low fraction BS (1)	High fraction of vegetation (2)	Low mean reflectance (3)	High unmixing RMSE (4)	Sum of cover fractions very low or high (5)
Northern section	0.35	0.20	0.25	0.34 ^a	10	8 ^b	0.4–1.4	32.5	47.5	13.6	5.4	0.7	0.4
Southern section	0.35	0.25	0.30	0.41 ^a	10	8 ^b	0.4–1.4	10.1	79.5	9.9	0.0	0.5	0.0
Entire dataset								26.8	55.6	12.7	4.0	0.6	0.3

^a Sum of PV and NPV * 0.75, ^b Average scene RMSE plus 8 times its standard deviation, ^c If more than one reason is eminent to remove the pixel, the pixel is flagged according to the first threshold that is failed (in order of the exclusion code number). Thresholds (left side) separate pixels that are suitable for calculating a residual soil spectrum. Total percentage of image pixels (right side), where a residual soil spectrum was calculated (valid pixels) and the percentages of pixels that are excluded for each factor. Each of them is given an exclusion code number that corresponds to the numbers in Fig. 3e.

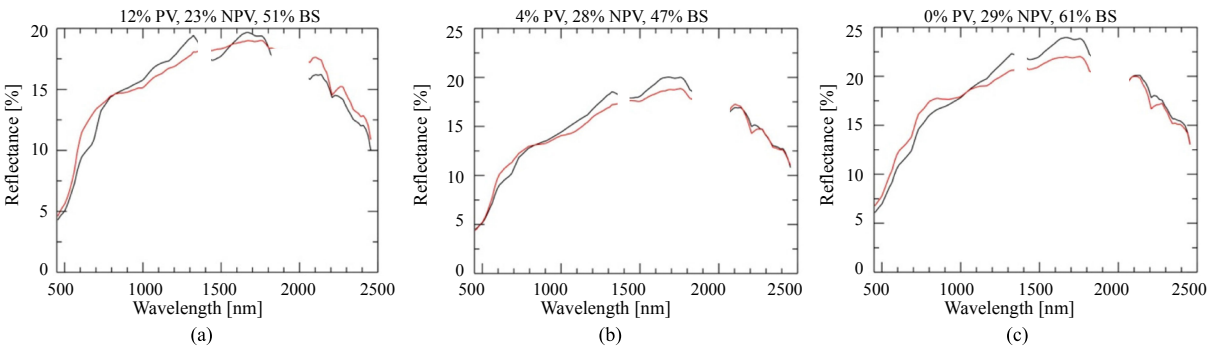


Fig. 5. Examples for pixel residual soil spectra calculated based on the hyperspectral imagery and derived ground cover information (red lines) compared to original HyMap pixel spectra (black lines). Major water bands are masked. Estimated cover fractions of green (PV) and dry vegetation (NPV) and bare soil (BS) are given for each location.

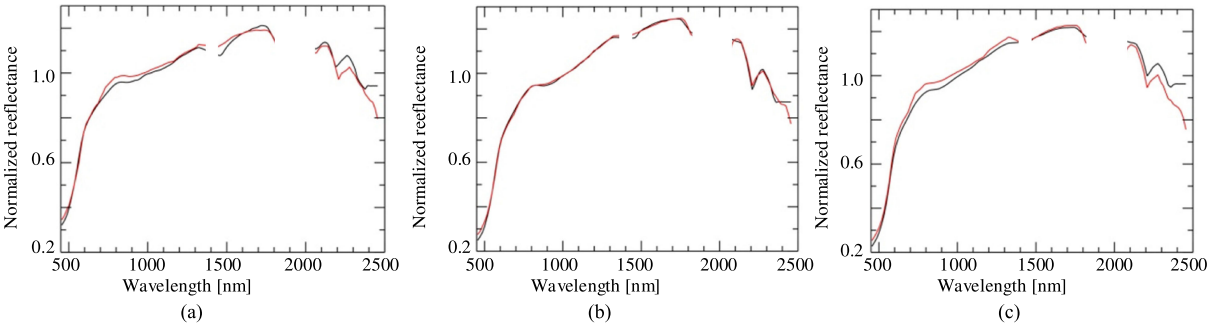


Fig. 6. Examples for residual soil spectra (red lines) compared to field reference spectra (black lines). For the pixel shown in (a), 1% NPV and 99% BS were calculated in spectral unmixing and the RMSE of the fit between the spectra is 0.029. (b) has 4% NPV and 96% BS with RMSE of 0.015, and (c) has 19% NPV and 81% BS with RMSE of 0.041. Major water bands are masked.

In areas where a mixture of the three land cover classes occurs, the vegetation signal was reduced well by the approach, and in particular in the VIS range, a significant reduction of the red edge was apparent [see Fig. 5(a) and (b)]. Only where the unmixing does not identify the green vegetation present in a pixel, the red edge remained as artifact in the spectra [see Fig. 5(c)]. Signatures of residual soil spectra show high correlation in

spectral shape with corresponding field reference spectra (see Fig. 6 for examples). RMSE is 0.048 for all plots where an RSS was calculated.

B. Performance of Feature-Based SOC Prediction Model

The method of feature-based regression is evaluated upon significance and consistency for three soil constituents,

TABLE IV
OVERVIEW OF STATISTICS OF PREDICTED SOC CONTENTS FOR THE ENTIRE
STUDY AREA

	% of SOC
Calibration range [% SOC] (c_{min} , c_{max})	0.21–5.85
Statistics of predicted SOC contents [% SOC]	Mean: 1.22 Stdev: 0.61 Min: 0.21 Max: 5.84
	% of pixels
Negative values calculated ($c_{calc} < 0$)	1.1
Calculated contents below calibration range ($0 < c_{calc} < c_{min}$)	3.1
Calculated contents within calibration range ($c_{min} \leq c_{calc} \leq c_{max}$)	95.8
Calculated contents exceed calibration range ($c_{calc} > c_{max}$)	0.00001

different measurement set-ups, and spectral resolutions in [29]. It was shown that the performance of this method lies within the range PLS regression techniques can provide, though the models are stronger linked to spectral features, which generally presents advantages in terms of significance and transferability (e.g., [19]).

Table II gives details of the applied SOC prediction model and of the spectral variables that dominate this model. The model applies 12 out of 16 spectral variables and variables describing the spectral hull in the VIS and SWIR range are of great importance (see Table II). The information given in Table II demonstrates the weak correlation between spectral variables as related variables (e.g., depth and area of an absorption feature) do not show similar influence in the regression relationship. The model provides good calibration accuracies with R_{Cal}^2/R_{Val}^2 of 0.81/0.62 with $RMSE_{Cal}/RMSE_{Val}$ about 0.43%. Accuracy in test set validation is slightly lower than calibration accuracy but still within the accuracy of medium class models [16]. With this, the accuracy of the SOC-prediction model used here is in the same range or slightly lower compared to studies in mainly agricultural environments (e.g., [4], [15]).

C. SOC Contents Predicted for the Albany Thicket Biome

The map of predicted SOC for the Albany Thicket Biome, South Africa, provides a snapshot for the spatial distribution in 2009 [see Fig. 2(d)]. For 96% of the pixels within the study area, for which residual soil spectra are calculated and SOC is predicted, the estimated SOC is within the calibration range of the regression model (see Table IV). This shows that the field sampling represented well the variability of the investigated parameters in areas for which soil constituents can be predicted based on hyperspectral imagery and that the prediction model is well adapted to the present conditions.

The average of predicted SOC contents for the entire image area, with 1.14% SOC is very close to the chemical reference of the 163 collected samples (see [29], Table IV). Only very low contents seem overestimated and very high contents seem underestimated, which is likely an effect of the spatial resolution of the HyMap sensor smoothing local extremes of limited spatial extent, or of the spatial distribution of collected samples.

Scatter plots of measured versus calculated SOC contents and the coefficient of determination (R^2) show no significant relationship for the regional validation of the multistage methodology in the South African study area [see Fig. 7(a)]. Modeling residues indicate a slight overestimation of actual SOC contents, with predicted SOC contents being on average 0.28% higher for the available samples [see Fig. 7(b)].

V. DISCUSSION

A. Spatial Extension of SOC Predictions

The percentage of pixels that are extracted as being suitable for a prediction of SOC with our methodology are compared to two previously applied approaches for this purpose (see Fig. 8 for local example). Fig. 8(a) is limited to BS surfaces (abundance of PV and NPV in spectral unmixing are 0) while Fig. 8(b) comprises pixels with up to 10% vegetation coverage (PV+NPV) (e.g., used in [24]). They include only 2.3 and 5.8% of the image pixels of the sample area, respectively. Consequently, no spatially continuous information can be provided, and SOC predictions from these data cover only local regions such as the prominent BS areas. In addition, SOC predictions from pixel spectra including any amounts of vegetation are biased by this, as prediction errors are expected to increase with the permitted amount of vegetation included in the analyzed spectral signature (compare, e.g., [26], [27]).

When applying the proposed methodology of resolving spectral mixtures and calculating residual soil spectra, the amount of pixels within the sample area is increased to 45.4% [see Fig. 8(c)]. These percentages of pixels are similar for the entire northern part of the study area, though less for the southern part due to the generally higher vegetation coverage. This shows that the presented calculation of residual soil spectra significantly increases the number of image pixels for which SOC contents subsequently can be predicted. Predicted soil constituents are further expected to be more accurate because the influence of vegetation on analyzed soil signatures is minimized.

B. Interpretation of Derived Spatial SOC Patterns

The spatial patterns of SOC in derived maps can be linked to geomorphic features and processes within the ecosystem which are supported by a number of field observations. SOC patterns (see Fig. 2(d) for overview and Fig. 3(g) for local example) are a result of the interdependence of the duration and intensity of the enrichment in SOC by the accumulation of plant litter and the reduction of SOC by soil erosion. This relationship is affected by topography (see, e.g., [20], [24]) and in the Albany Thicket Biome is significantly modified by the nonhomogenous occurrence of dense pristine vegetation and open-degraded vegetation. Scatter plots reveal an existing, however, not particularly strong correlation between SOC contents and the two vegetation types (see [39] for details).

Based on the relation between carbon input and erosion, the ecosystem can be divided in two zones (after [49]). Soil development is progressive under relative stability, where soils become

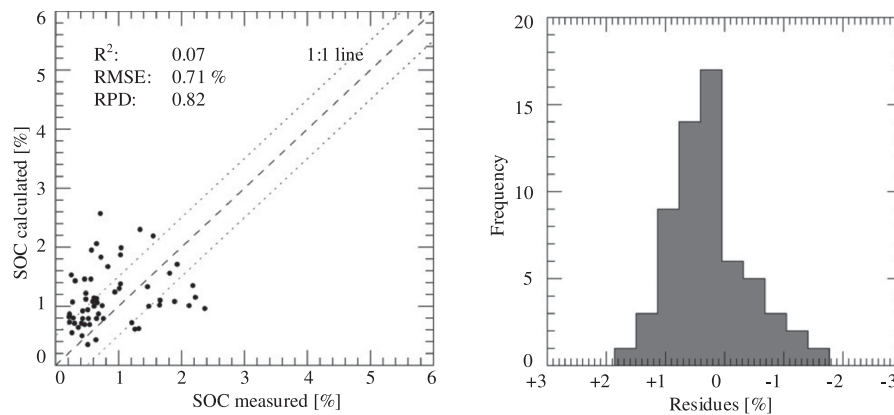


Fig. 7. Regional validation of the multistage methodology. Scatter plots for *in situ* measured versus predicted SOC (a) and histograms of modeling residues as the deviation between modeled and *in-situ* measured contents constituent (b). In (a) the deviation of 0.5% as the assumed accuracy of the laboratory method is highlighted.

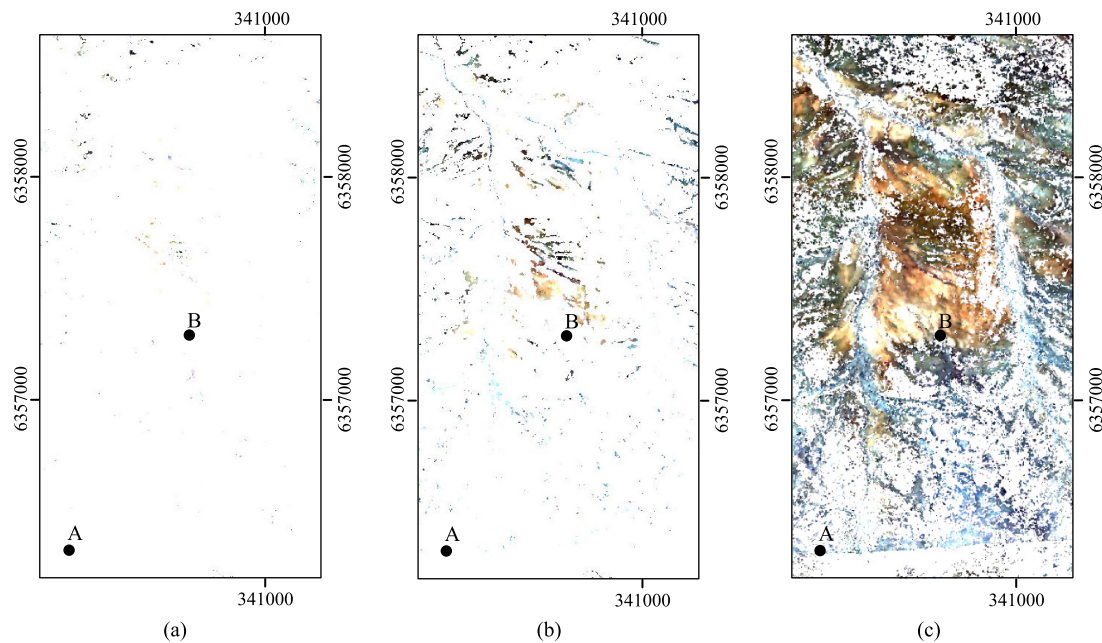


Fig. 8. True color images of data sets that can be used for the prediction of SOC for a sample area of about 4 km² in the Albany Thicket Biome, South Africa: (a) includes only image pixels comprising bare soil surfaces, (b) image pixels with up to 10% vegetation cover, and (c) residual soil spectra calculated using the proposed methodology. See Fig. 3(c) for true-color image of this subset.

more developed and the topsoil layer is enriched in organic carbon. In the study area, this occurs on partly vegetated flat areas or gently dipping slopes. In addition to being the provider of organic material, the vegetation prevents the organic rich topsoil layer from being eroded [50]. This is reflected in deeper soil profiles observed in the field (not shown here). A local example is shown in location A in Fig. 3 on the flat areas surrounding the bare plain in the center and the slopes gently rising to the south (SOC about 1.5%). It is expected that SOC contents are even higher in densely vegetated areas [33], though this information cannot be derived from the soil maps because of the masking effect of vegetation. The positive effects of high concentrations of organic carbon on soils making these soils more resistant to soil erosion is especially important in semiarid areas to sustain soil fertility and mitigate droughts [51].

Despite this, soil development is regressive and SOC contents are low where active erosion reduces the amount of newly built soil material more than it was accumulated. This applies for three distinctive regions within the study area: 1) Bare areas where almost no vegetation is present, 2) Elevated ridges and slopes with medium to high gradients, where a reduced vegetation cover is unable to prevent the erosion of the organic topsoil layer, 3) Larger runoff trenches that are not overgrown by vegetation. A typical example is found on the flat bare plain in the center of Fig. 3 (location B, SOC mostly below 0.6%). Low SOC contents account for limited plant growth and make soils prone to erosion due to a lowered structure stability. In the field, the lower vegetation cover was found and gullies were observed in such areas (data not shown). Soil profile depth was often found reduced and the exposure of sandy bedrocks was observed frequently, indi-

cating long-term soil erosion. The bare plain in Fig. 3 is an example for a degradation hot spot that is identified upon reduced vegetation coverage and low carbon contents based on image data which is supported by indicators for active soil erosion visible in the field.

SOC contents in the southern section on average are 0.8% SOC higher than in the northern section of the study area [see Fig. 2(d)]. The reason for this is found in the generally high vegetation coverage due to higher precipitation amounts and low relief in the southern section. This is corroborated by soil profiles consistently being deeper in the southern transect which suggests a higher stage of soil development (data taken in 2011, see [39]).

C. Limitations of SOC Predictions in the Semiarid Study Area

The lower accuracy of the SOC-prediction model in comparison to other studies done in agricultural environments and the poor validation performance of the regional validation of the entire methodology are likely a result of the large size and the highly variant characteristics of the South African study area. Two main reasons are identified.

First, the spectral mixtures of surface materials, which in this nonagricultural environment occur on a very small scale and are complex due to numerous surface materials (variability in vegetation, geology, soil types, etc.). Already previous studies reported that the accuracy of soil constituents' predictions declines with the intensity of spectral mixture within pixels. For instance, this was found by [26] who predicted SOC contents for agricultural maize fields by applying a similar technique of removing vegetation spectral influence and by [27], who used blind source separation to resolve spectral mixtures and to predict clay contents in vineyard regions. Yet, the importance of resolving spectral mixtures was apparent in both agricultural environments, where spectral mixture between soil and vegetation for maize fields and vineyards was limited to one vegetation type, and thus, was well identifiable and resolvable. Also, [14], who used Hyperion satellite data for SOC predictions in Australian soils, attributed low validation correlations besides to the high noise present in Hyperion spectra to the intense spectral mixing within pixels (30-m spatial resolution). Improvements made by our approach are shown by the qualitative analysis of the residual soil spectra and their comparison to field spectra. Further improvements could be implemented by applying higher spatial resolution sensors, and thereby decreasing the amount of spectral mixtures in pixel spectra and improving their resolution by spectral unmixing, or by enhancing the accuracy in geo-correction, which would allow for a better correlation of field control and airborne measurements.

A second reason limiting the accuracy of soil constituents predictions are the generally low values and the small variability in measured SOC contents (see Fig. 3(a) in [29]) in comparison to the assumed accuracy of the laboratory method of 0.5% SOC (e.g., [15], [47]). This is especially a factor lowering prediction RPD of the SOC model, corresponding to findings reported in, e.g., [15] and [47]. This issue could only be addressed by improved wet chemical analysis reducing measurement er-

rors, and thereby, allowing for a possibly more accurate model calibration.

In addition, the geometric accuracy of the data with typical unsystematic errors in position of about two to three pixels also has an influence on the validation with ground data, which is increasingly important in patchy landscapes. This could be improved wherever high-quality ground reference data and digital elevation models are available, however, this might not be the case for wide landscapes that are for instance hot spots for degradation or carbon sequestration. Differences in soil-water content between field samples and image data acquisition can be neglected as a reason for the poor validation performance. In 2009, stable dry weather conditions have caused minimal changes in vegetation coverage and soil moisture, with weeks of no rain before the field sampling and hyperspectral image acquisition (data by [52]). Prediction residues also seem not to be linked with the fractional cover of the two vegetation types (see [39], for details).

VI. CONCLUSION

In semiarid regions, which are especially vulnerable to climate change and human-induced land-use changes, large-area SOC mapping using IS is complex due to locally varying land surface and soil conditions, the partial vegetation coverage, and usually low soil carbon contents. So far no consistent remote sensing method is available that accounts for all disturbances and influences introduced by this natural environment. This study addressed these difficulties by extending existing approaches leading towards the mapping of SOC contents to facilitate their application in semiarid environments and tested the multistage methodology in the Albany Thicket Biome in South Africa.

Results showed that the proposed methodology supported the prediction of SOC in such areas in several ways. A calculation of residual soil spectra from mixed HyMap pixels before large-area SOC prediction allowed to significantly reduce the influence of other materials such as vegetation on spectral signatures of the hyperspectral imagery. Thereby, SOC could be predicted for a much larger area compared to previously used approaches, where analyzed pixels were limited to ones with no or a very low vegetation coverage (increase from <6 to 45% of the study area). This allowed the provision of soil information nearly spatially continuously for soil dominated areas, including information from smaller patches with a reduced vegetation cover (up to 40% green and dry vegetation combined) in areas dominated by vegetation. Using knowledge from field campaigns, derived spatial patterns of SOC were interpreted and linked to geomorphic features and long-term processes within the study area, such as ongoing erosion. It, thus, permitted a temporal perspective that goes beyond the sole analysis of current vegetation coverage at one point in time.

On the other hand, when applying accepted validation standards, a quantitative prediction of SOC contents could not be achieved for the entire study area of 320 km². The accumulated uncertainties of the multistage method (including spectral unmixing and soil signature reconstruction, wet chemistry

analysis for low SOC contents, regression modeling with a limited sample variety) were too large in relation to the generally low SOC content in this nonagricultural semiarid environment. This shows that factors remain above the level that is accounted for in this study that disturb the accurate prediction of SOC.

Additional processing steps have to be included to provide a methodology that facilitates a quantitative prediction of SOC in areas aside of agricultural regions including the complex conditions in semiarid areas. This could, e.g., include the regionalization of the prediction models, or improvements within the spectral unmixing model. The focus of further research should be drawn to this matter, considering the vulnerability of semiarid areas to climatic and land-use changes. Its importance rises in particular when envisioning the application of future hyperspectral satellite sensor systems such as EnMAP, HISUI, and HypIRI for the mapping of key soil constituents in extensively used semiarid ecosystems. These sensors provide the capacity and technical requirements for determining the spatial distribution of various ecosystem parameters over large areas and in multitemporal surveys, but need accurately calibrated models in combination with adequate approaches to resolve spectral mixtures.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers to help improving and focusing the manuscript.

REFERENCES

- [1] A. Mills *et al.*, "Ecosystem carbon storage under different land uses in three semi-arid shrublands and a mesic grassland in South Africa," *South African J. Plant Soil*, vol. 22, pp. 183–190, 2005.
- [2] A. Don, J. Schumacher, and A. Freibauer, "Impact of tropical land-use change on soil organic carbon stocks—A meta-analysis," *Global Change Biol.*, vol. 17, pp. 1658–1670, 2011, doi: 10.1111/j.1365-2486.2010.02336.x.
- [3] E. Ben-Dor, K. Patkin, A. Banin, and A. Karnieli, "Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—A case study over clayey soils in Israel," *Int. J. Remote Sens.*, vol. 23, pp. 1043–1062, 2002.
- [4] A. Stevens *et al.*, "Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils," *Geoderma*, vol. 144, pp. 395–404, 2008.
- [5] R. Viscarra Rossel, D. Walvoort, A. McBratney, L. Janik and J. Skjemstad, "Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties," *Geoderma*, vol. 131, pp. 59–75, 2006.
- [6] IPCC, "Climate Change 2007 – Synthesis Report," Report at IPCC Plenary XXVII, Valencia, Spain, 2007.
- [7] E. Ben-Dor and A. Banin, "Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils," *Remote Sens. Environ.*, vol. 48, pp. 261–274, 1994.
- [8] E. Ben-Dor and A. Banin, "Near infrared analysis (NIRA) as a simultaneously method to evaluate spectral featureless constituents in soils," *Soil Science*, vol. 159, pp. 259–269, 1995.
- [9] J. Dematté and G. Gracia, "Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance," *Soil Sci. Soc. Amer. J.*, vol. 63, pp. 327–342, 1999.
- [10] M. Nocita *et al.*, "Soil spectroscopy: An alternative to wet chemistry for soil monitoring," *Adv. Agronomy*, vol. 132, pp. 139–159, 2015.
- [11] H. Bartholomeus *et al.*, "Spectral reflectance based indices for soil organic carbon quantification," *Geoderma*, vol. 145, pp. 28–36, 2008.
- [12] E. Ben-Dor, Y. Inbar, and Y. Chen, "The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process," *Remote Sens. Environ.*, vol. 61, pp. 1–15, 1997.
- [13] C. Gomez, P. Lagacherie, and G. Coulouma, "Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements," *Geoderma*, vol. 148, pp. 141–148, 2008.
- [14] C. Gomez, R. Viscarra Rossel, and A. McBratney, "Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study," *Geoderma*, vol. 146, pp. 403–411, 2008.
- [15] A. Stevens *et al.*, "Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy," *Geoderma*, vol. 158, pp. 32–45, 2010.
- [16] C.-W. Chang, D. Laird, M. Mausbach, and C. R. Hurburgh Jr., "Near-infrared reflectance spectroscopy—Principal components regression analyses of soil properties," *Soil Sci. Soc. Amer. J.*, vol. 65, pp. 480–490, 2001.
- [17] D. J. Brown, R. S. Bricklemyer, and P. R. Miller, "Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana," *Geoderma*, vol. 129, pp. 251–267, 2005.
- [18] A. Stevens, B. van Wesemael, G. Vanderschrick, S. Touré, and B. Tychon, "Detection of carbon stock change in agricultural soils using spectroscopic techniques," *Soil Sci. Soc. Amer. J.*, vol. 70, pp. 844–850, 2006.
- [19] V. Mulder, S. de Bruin, M. Schaepman, and T. Mayr, "The use of remote sensing in soil and terrain mapping—A review," *Geoderma*, vol. 162, pp. 1–19, 2011.
- [20] J. Hill and B. Schütt, "Mapping complex patterns of erosion and stability in dry mediterranean ecosystems," *Remote Sens. Environ.*, vol. 74, pp. 557–569, 2000.
- [21] E. Ben-Dor *et al.*, "Quantitative mapping of the soil rubification process on sand dunes using an airborne hyperspectral sensor," *Geoderma*, vol. 131, pp. 1–21, 2006.
- [22] S. Chabrilat, A. Goetz, L. Krosleyand and H. Olsen, "Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution," *Remote Sens. Environ.*, vol. 82, pp. 431–445, 2002.
- [23] P. Lagacherie, F. Baret, J.-B. Feret, J. Madeira Netto, and J. Robbez-Masson, "Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements," *Remote Sens. Environ.*, vol. 112, pp. 825–835, 2008.
- [24] W. Schwanghart and T. Jarmer, "Linking spatial patterns of soil organic carbon to topography—A case study from south-eastern Spain," *Geomorphology*, vol. 126, pp. 252–263, 2011.
- [25] C. Elvidge, "Visible and near infrared reflectance characteristics of dry plant materials," *Int. J. Remote Sens.*, vol. 11, pp. 1775–1795, 1990.
- [26] H. Bartholomeus *et al.*, "Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy," *Int. J. App. Earth Observation Geoinf.*, vol. 13, pp. 81–88, 2011.
- [27] W. Ouerghemmi, C. Gomez, S. Naceur, and P. Lagacherie, "Applying blind source separation on hyperspectral data for clay content estimation over partially vegetated surfaces," *Geoderma*, vol. 163, pp. 227–237, 2011.
- [28] M. H. D. Franceschini *et al.*, "Prediction of soil properties using imaging spectroscopy: Considering fractional vegetation cover to improve accuracy," *Int. J. Appl. Earth Observation Geoinf.*, vol. 38, pp. 358–370, 2015.
- [29] A. Bayer, M. Bachmann, A. Müller, and H. Kaufmann, "A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African ecosystem," *Appl. Environ. Soil Sci.*, vol. 2012, 2012, Art. no. 971252, doi: 10.1155/2012/971252.
- [30] D. Hoare *et al.*, "Albany thicket biome," in *The Vegetation of South Africa, Lesotho and Swaziland*, L. Mucina and M. C. Rutherford, Eds. Pretoria, South Africa: South African Nat. Biodiversity Inst., 2006, pp. 541–567.
- [31] R. Lechmere-Oertel, G. Kerley, and R. Cowling, "Patterns and implications of transformation in semi-arid succulent thicket, South Africa," *J. Arid Environ.*, vol. 62, pp. 459–474, 2005.
- [32] J. Lloyd, E. van den Berg, and A. Palmer, "Patterns of transformation and degradation in the thicket biome, South Africa," *Terrestrial Ecol. Res. Unit, Univ. of Port Elizabeth, Port Elizabeth, South Africa, Tech. Rep.* 39, 2002.
- [33] A. Mills and M. Fey, "Transformation of thicket to savanna reduces soil quality in the Eastern Cape, South Africa," *Plant Soil*, vol. 265, pp. 153–163, 2004.

- [34] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, "The HyMap airborne hyperspectral sensor: The system, calibration and performance," presented at the 1st EARSEL Workshop Imaging Spectroscopy, Zurich, Switzerland, 1998.
- [35] D. Schl  fer and R. Richter, "Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: Parametric orthorectification," *Int. J. Remote Sens.*, vol. 23, pp. 2609–2630, 2002.
- [36] R. Richter, "Atmospheric/topographic correction for airborne imagery," German Aerospace Center, Oberpfaffenhofen, Germany, Tech. Rep. DLR-IB 565-02/10, 2010.
- [37] D. Rogge, M. Bachmann, B. Rivard, and J. Feng, "Hyperspectral flight-line leveling and scattering correction for image mosaics," in *Proc. Int. Geosci. Remote Sens. Symp.*, Munich, Germany, 2012, pp. 4094–4097.
- [38] D. Rogge and B. Rivard, "Iterative spatial filtering for reducing intra-class variability and noise," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Reykjavik, Iceland, 2010, pp. 1–4.
- [39] A. Bayer, "Methodological developments for mapping soil constituents using imaging spectroscopy," Ph.D. dissertation, Universit  t Potsdam, Potsdam, Germany, 2013.
- [40] R. N. Feudale *et al.*, "Transfer of multivariate calibration models: A review," *Chemometrics Intell. Lab. Syst.*, vol. 64, pp. 181–192, 2002.
- [41] C. Hecker, "Mapping feldspars from above—A thermal infrared and partial least squares-based approach," Ph.D. dissertation, University of Twente, Enschede, The Netherlands, 2012.
- [42] T. Jarmer, "Spectroscopy and hyperspectral imagery for monitoring summer barley," *Int. J. Remote Sens.*, vol. 34, pp. 6067–6078, 2013.
- [43] D. Rogge *et al.*, "Integration of spatial–spectral information for the improved extraction of end members," *Remote Sens. Environ.*, vol. 110, pp. 287–303, 2007.
- [44] J. Adams, M. Smith, and P. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res.*, vol. 91, pp. 8098–8112, 1986.
- [45] M. Bachmann, "Automatisierte ableitung von bodenbedeckungsgraden durch MESMA-Entmischung," Ph.D. dissertation, Julius-Maximilians-Universit  t W  rzburg, W  rzburg, Germany, 2007.
- [46] D. Rogge, B. Rivard, J. Zhang, and J. Feng, "Iterative spectral unmixing for optimizing per-pixel endmember sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 12, pp. 3725–3736, Dec. 2006.
- [47] T. Udelhoven, C. Emmerling, and T. Jarmer, "Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study," *Plant Soil*, vol. 251, pp. 319–329, 2003.
- [48] K. McGwire, T. Minor, and L. Fenstermaker, "Hyper-spectral mixture modeling for quantifying sparse vegetation cover in arid environments," *Remote Sens. Environ.*, vol. 72, pp. 360–374, 2000.
- [49] P. Birkeland, "Soil-geomorphic research—A selective overview," *Geomorphology*, vol. 3, pp. 207–224, 1990.
- [50] M. A. Weltz, M. R. Kidwell, and H. D. Fox, "Influence of abiotic and biotic factors in measuring and modeling soil erosion on rangelands: State of knowledge," *J. Range Manage.*, vol. 51, pp. 482–495, 1998.
- [51] H. Tiessen, E. Cuevas, and P. Chacon, "The role of soil organic matter in sustaining soil fertility," *Nature*, vol. 371, pp. 783–785, 1994.
- [52] Agricultural Research Council, Institute for Soil, Climate and Water (ARC-ISCW), Provision of climate station data, 2011.



Anita D. Bayer has a background in technical geosciences and received the Ph.D. degree in remote sensing from the University of Potsdam, Potsdam, Germany, in 2013.

From 2009 to 2013, she was working with the Applied Spectroscopy Group at the German Remote Sensing Data Center, German Aerospace Center (DLR), Wessling, Germany, on image-based techniques for soil parameter quantifications in semiarid natural areas. Her research interests include the detection and quantification of changes in ecosystem

function and ecosystem services using remote sensing and modeling techniques. Since 2013, she has been working on exploring ecosystem function and services among others in response to climate and land-use change using global vegetation modeling at Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany.



Martin Bachmann received the Diplom-Geogr. degree from the Technical University of Dresden, Dresden, Germany, in 2002, and the Ph.D. (Dr. rer. nat.) degree from the University of W  rzburg, W  rzburg, Germany, in 2007. His dissertation was on automated estimation of ground cover fractions using multiple endmember unmixing for dryland applications.

Since 2005, he has been a Research Scientist with the Applied Spectroscopy Group, German Remote Sensing Data Center, German Aerospace Center (DLR), Wessling, Germany, leading the team between 2009 and 2013. For eight years, he was in charge of the data preprocessing chain of DLR's Optical Airborne Remote Sensing and Calibration Facility (OpAIRS). He leads the work package on "Data Quality Control" within the EnMAP ground segment, and is involved in the FP7 EUFAR Joint Research Activity on Data Quality Control for airborne hyperspectral and LiDAR preprocessing. His research interests include the development and application of spectral unmixing approaches, and algorithm developments toward the derivation of soil parameters using field, laboratory, and airborne spectroscopy.



Derek Rogge received the Ph.D. degree at the Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Canada, in 2007.

He is currently a Research Scientist with the Applied Spectroscopy Group, German Remote Sensing Data Center, German Aerospace Center (DLR), Wessling, Germany. His present research interests include algorithm development related to endmember extraction, spectral unmixing, and their implementation with large multi flight-line hyperspectral datasets, simulated EnMAP, and Sentinel 2 data. His applied remote sensing interests include geological and soil applications, such as regional mapping, target mapping for mineral exploration, mine waste, environmental monitoring, and fractional cover estimates for soil erosion modeling.



Andreas M  ller received the Diploma degree from the Working-Group for Geo-Scientific Remote Sensing, Munich University, Munich, Germany, in geology and remote sensing in May 1990.

He was the Head of Department "Land Surface" at the German Remote Sensing Data Center, German Aerospace Center. He has 25 years of experience in the processing and evaluation of airborne and spaceborne imaging spectrometer and multispectral data, calibration/validation, atmospheric/geometric correction, and information extraction. He was leading spectral ground validation campaigns in Germany, other European countries, Israel, and India. He is currently the Coinvestigator on the German hyperspectral satellite mission Environmental Mapping and Analysis Program (EnMAP) and a Member of the EnMAP Science Advisory Board. Since 1996, he has been the Chairman of the Special Interest Group "Imaging Spectroscopy" at the European Association of Remote Sensing Laboratories. He currently coordinates the land surface remote sensing activities at the German Remote Sensing Data Center with specific focus on multisensor information extraction in support of applications in the fields of land surface processes and interactions, management of natural resources, and urban development.



Hermann Kaufmann received the Dr. rer. nat. degree from the Munich Ludwig-Maximilians University, Munich, Germany, in geology and remote sensing. In 1992, he received the Dr. Ing. Habil degree in the field of remote sensing at the Faculty of Engineering, University of Karlsruhe.

He was the Head of Department 1—Geodesy and Remote Sensing, GFZ German Research Centre for Geosciences, Potsdam, Germany, and held a Chair at the University of Potsdam. His major scientific experiences include the fields of data processing, sensor definition, and applications dedicated to various disciplines. He was the Scientific Leader of the EnMap Hyperspectral Satellite Program, to be launched in 2018.